Minireview

# Comparative genomics of *Arabidopsis* and maize: prospects and limitations

Volker Brendel*, Stefan Kurtz† and Virginia Walbot‡

Addresses: *Department of Zoology and Genetics and Department of Statistics, Iowa State University, Ames, IA 50010, USA. †Technische Fakultät, Universität Bielefeld, D-33501 Bielefeld, Germany. ‡Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020, USA.

Correspondence: Volker Brendel. E-mail: vbrendel@iastate.edu

Published: 14 February 2002

## Abstract

The completed *Arabidopsis* genome seems to be of limited value as a model for maize genomics. In addition to the expansion of repetitive sequences in maize and the lack of genomic micro-colinearity, maize-specific or highly-diverged proteins contribute to a predicted maize proteome of about 50,000 proteins, twice the size of that of *Arabidopsis*.

Maize (*Zea mays* L., corn) was domesticated in the highlands of Central Mexico approximately 10,000 years ago [1]. Corn agriculture spread rapidly into diverse climate zones, ranging from 45° N to 45° S, and supported vast Native American civilizations. Today, maize is one of the world's most important crops: for direct human consumption, as a key component of animal feed, and as the source of chemical feed stocks. Grass species (including maize) cover 20% of the terrestrial surface of the earth, and the grains from maize, rice, wheat, and minor grass crops provide the majority of calories in the human diet [2].

Since the beginning of the twentieth century, maize has been a model species for genetic analysis, reflecting its unusual biological features. Maize plants produce separate male and female inflorescences, which greatly facilitates experimentally controlled pollination by eliminating the need for emasculation (Figure 1). Large numbers of progeny (300-600 kernels per ear) and the ease of crossing allow a single maize geneticist to generate more than 100,000 outcross progeny per day. Individual plants produce up to $10^7$ pollen grains, allowing fine-structural genetic mapping for phenotypes that can be scored at the pollen stage. Using this abundant material and extraordinary natural diversity, early geneticists mapped many genes, uncovered subtle genetic phenomena such as paramutation and imprinting, and made practical contributions to agriculture through the discovery of hybrid vigor and cytoplasmic male sterility.

The beautiful detail evident in meiotic maize chromosomes stimulated a generation of gifted cytogeneticists to identify the physical basis for recombination, to construct linkage maps tied to chromosomes, and to analyze the consequences of chromosome breakage. Of particular importance to current functional genomics was Barbara McClintock's discovery of transposable elements by analyzing the regulation of somatic variegation and germinal mutation in maize. Once maize transposons were molecularly cloned, they provided the means to clone any tagged gene: maize provided the first discovery of many plant-specific gene products and facilitated the cloning of related genes from other flowering plants. The availability of detailed genetic knowledge, a large community of researchers, and ease of gene cloning and genetic analysis make maize the monocotyledenous species of choice for many studies.

The maize genome is organized into 10 chromosomes (2N = 20), and is about 2.4 x $10^9$ base-pairs in total. Sorghum, which is estimated to have diverged from a common ancestor with maize about 15-20 million years ago (MYA), has the same chromosome number, but its genome is about one third of the size. Rice diverged from a common

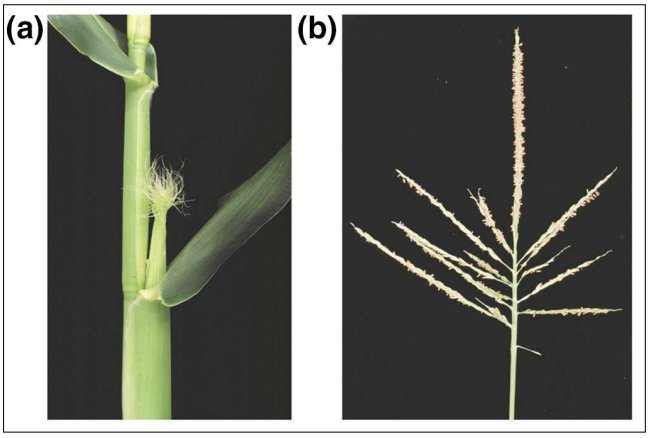


**Figure 1**
Maize inflorescences. The separation of **(a)** female inflorescence (ear) and **(b)** male inflorescence (tassel) is one of the key features of the maize plant responsible for its pivotal role in plant genetics, greatly simplifying controlled pollination (photos courtesy of Tom Peterson, Iowa State University).

ancestor with maize and sorghum about 50-60 MYA and has 12 chromosomes (2N = 24), comprising a much smaller genome of about 430 million base-pairs. Comparative genomics of these grasses suggests considerable colinearity between their genomes [3]. The size differences of the genomes are presumed to result from the ancestral allotetraploidization (approximate duplication from diploid to tetraploid when two species hybridize) of the maize genome [4] and differences in the expansion and dispersion of repetitive DNA (long terminal repeat retrotransposons, miniature inverted repeat transposons, and other repetitive sequences) [5].

In December 2000, *Arabidopsis thaliana* became the first plant species for which the genome was almost entirely sequenced (currently, 117 of an estimated 125 million base-pairs are available, with only centromeric and ribosomal DNA repeat regions as yet unsequenced [6]; reviewed in [7]). Because of its small genome size, ease of transformation, and tolerance of life in a growth chamber, this seemingly lowly weed has emerged as the model flowering plant, ahead of commercially important crops. The choice will be well justified if the evolutionarily recent advent of flowering plants means that most genes found in *Arabidopsis* prove to be common to all flowering plants. Among the crops, members of the *Brassica* genus (including *B. oleracea* and *B. rapa*, the so-called 'cole-crops', oilseeds, and mustard) are most closely related to *Arabidopsis* (divergence less than 20 MYA). Gene order seems to be largely conserved, and thus the *Arabidopsis* genome should prove a powerful tool for studying *Brassica* genomics [8,9]. Significant colinearity has also been observed between *Arabidopsis* and soybean [10] (divergence time 100 MYA), and *Arabidopsis* and tomato [11,12] (divergence time more than 100 MYA). This article assesses the prospects for comparative maize-*Arabidopsis* genome analysis in view of the greater divergence time (more than 150 MYA) between grasses (which are monocots) and flowering plants (dicots).

## Lack of synteny between maize and *Arabidopsis*

The extent of conservation of gene order between the grasses and *Arabidopsis* can be estimated from three well-studied groups of maize loci: the *a1-sh2* region [13-15], the *adh1* region [16,17], and the *bz* locus and its associated genes [18]. The *a1-sh2* region in maize, sorghum, and rice contains the *sh2* gene upstream of *a1*, transcribed in the same direction. The *a1* gene encodes an NADPH dihydroflavonol reductase required for anthocyanin biosynthesis and *sh2* encodes an endosperm-expressed ADP glucose pyrophosphorylase important in starch biosynthesis. The two genes are separated by about 140 kilobases (kb) in maize but only about 19 kb in sorghum and rice. Moreover, *a1* is duplicated in sorghum. Sequences that are highly similar to *sh2* can be found on *Arabidopsis* chromosomes 1, 2, 4, and 5. Potential homologs of *a1* map to *Arabidopsis* chromosomes 2 and 5, but they are far apart from the potential *sh2* genes. Recently, two additional genes have been identified in the *a1-sh2* interval: *x1* and *yz1*, which are of unknown function and conserved among maize, rice, and sorghum [14,19].

Genic regions are generally conserved between the *adh1* regions of maize and sorghum, although *adh1* is the only gene with assigned function (alcohol dehydrogenase), and maize is missing three out of ten other potential genes within this region [16]. Whereas the maize region is replete with retrotransposons, gathered into sequence blocks of 14-70 kb and inserted between the potential genes, the sorghum sequence does not contain any retrotransposons. Colinearity with *Arabidopsis* appears limited to a block of two genes conserved between sorghum and *Arabidopsis* [16]. Interestingly, the colinearity of this locus pair is interrupted even between maize and rice [17].

The recently sequenced *bz* locus of maize and its chromosomal region displays a gene-dense genomic organization very different from *adh1*, with ten putative genes within a 32 kb stretch that is free of retrotransposons [18]. Although this gene density is similar to that in *Arabidopsis*, and most of the genes have potential homologs in *Arabidopsis* according to the genome sequence, no colinearity is evident. Thus, on the basis of our current picture of plant genome organization, micro-colinearity between different genomes may be even more limited than has previously been stated [20].

## Proteome comparisons

Although gene order does not appear to be conserved across the monocot-dicot divide, the repertoires of gene products (that is, the typical monocot and dicot proteomes) may be

conserved. This hypothesis cannot be fully tested until the complete *Arabidopsis* genome is matched to a complete monocot genome, but the current collection of maize proteins and genome sequence fragments may provide a clue. We downloaded the entire set of 4,195 maize protein sequence records from GenBank and reduced this collection to a representative, non-redundant set of maize proteins in several steps: firstly, removal of sequences less than 60 amino acids; secondly, removal of organelle-encoded proteins; and thirdly, selection of a single sequence to represent clusters of highly similar entries (including identities resulting from duplications in GenBank; this was done using the novel fast string matching program 'vmatch' [21]; V.B. and S.K., unpublished). The resulting set of 1,143 sequences was compared with a set of 25,617 putative *Arabidopsis* proteins [22] using BLASTP [23] at moderate stringency (BLAST -e option set to 1e-5). Most of the 117 entries without significant hits were identified as polypeptides encoded by transposable elements. The remaining sequences were matched directly against the *Arabidopsis* genome using the GeneSeqer spliced alignment program [24] to check for possible gene products not included in the *Arabidopsis* predicted protein set (only one unannotated *Arabidopsis* homolog of a maize protein was identified in this way). About 50 candidate maize-specific proteins remained, including several zeins, some predicted products of unknown function, and several other proteins (the latter group are listed in Table 1). On the basis of these results, we can give an upper estimate of 90% of maize proteins that have close homologs in *Arabidopsis*. The distinct maize genes appear to be tissue-specific (endosperm) or involved in maize pathogen-defense responses.

## Maize EST analysis

One pivotal strategy for identifying gene products involves sequencing of large sets of expressed sequence tags (ESTs). Many plant genome projects have adopted this approach, and there are currently more than 100,000 EST database entries in the public domain for each of soybean, tomato, *Medicago truncatula*, maize, *Arabidopsis*, and rice [25]. To further assess the overlap between the maize and *Arabidopsis* proteomes, we derived a set of 27,294 maize ESTs with non-redundant open reading frames (ORFs) of at least 120 codons (again using vmatch). The translated ORFs (derived from all six reading frames) were compared to the set of putative *Arabidopsis* proteins using BLASTP at different

**Table 1**

**Maize proteins with no obvious homologs in *Arabidopsis***

| Protein | GenBank accession number | Function |
|---|---|---|
| BETL(2-4) | CAB4466(2-4) | Anti-microbial, endosperm |
| Ribosome-inactivating proteins | S11859, CAC16167, P10593, T03942 | Anti-microbial, anti-fungal |
| Female gametophyte-specific protein ES3 | AAK08134 | Defensin |
| Basal layer anti-fungal peptides | CAC21604, CAC21605, CAC21607 | |
| Trypsin inhibitor | TIZM, TIZM1, S36236 | Anti-insect |
| RAB-17 | S08633 | Vesicle traffic |
| FDR3 | AAK53546 | Iron stress |
| ZmGR2(b,c) | BAA7480(6,7) | Gibberellin-responsive |
| Aluminum-induced proteins | AAB86493, T01322 | |
| ABA- and ripening-inducible-like protein | T02081 | |
| Bundle-sheath cell specific protein 1 | BAB20906 | C4 photosynthesis |
| Peroxidase K | AAC79955 | |
| Phytase | T04130 | Degradation of phytic acid, the main phosphor storage in maize seeds |
| ESR1c1 | CAA67122 | Endosperm-specific |
| Teosinte-branched protein 1 | AAK30124 | Associated with maize domestication (specific alleles) |
| Globulin 10 | C53234 | Storage |
| Ae(1,3) | CAB5655(2,3) | Amylase extender; modification of kernel starch composition |
| Arabinogalactan protein | AAF43497 | Cell-wall component |
| Probable membrane protein DAD1 | T01578 | |

The maize proteins were compared against the *Arabidopsis* protein set using BLASTP (see text for details). Maize query sequences are listed that did not match any *Arabidopsis* sequences at the 1e-5 level. Not listed: zeins, some sequences of highly biased composition and putative maize-specific proteins. Brackets ( ) are used to show related entries; for example, BETL2 has GenBank accession number CAB44662, and BETL3 CAB44663.

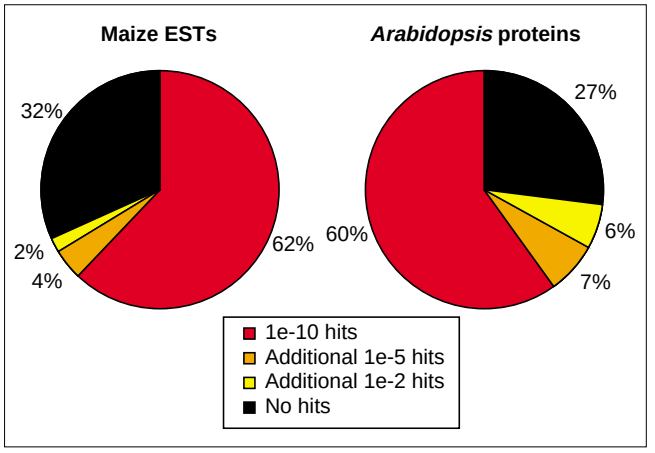


**Figure 2**
Comparison of maize proteins predicted from EST sequences with *Arabidopsis* proteins. A non-redundant set of protein sequences consisting of at least 120 amino acids each, derived from 27,294 distinct maize ESTs, was compared with 25,617 putative *Arabidopsis* proteins at different BLASTP stringency levels. The percentages in each pie chart give the fractions of the two sequence sets involved in these matches, at each stringency level.

stringency levels. As shown in Figure 2, 62-68% of the maize ESTs relate to ORF products that match *Arabidopsis* proteins, and the total fraction of the *Arabidopsis* protein set matched by the maize ESTs is 60-73%. Similar numbers were obtained for consensus sequences built from maize EST clusters [26]. Thus, a significant proportion of maize ESTs might encode highly diverged or maize-specific proteins. Some ORF products might not correspond to functional proteins, and incorrect gene prediction models and the as yet partial *Arabidopsis* protein set may also contribute to incomplete matching. For comparison, the same procedure applied to the *Arabidopsis* EST set compared to the *Arabidopsis* protein set gave a matching fraction of 88% or more of 28,161 qualifying ESTs, showing that chance ORFs may account for up to 12% of the unmatched ESTs in *Arabidopsis*, and presumably also in maize. We can therefore refine the estimate of maize proteins with close homologs in *Arabidopsis* to 60-90% of the maize proteome. Because ESTs are difficult to derive from genes expressed at low level there may in fact be more unmatched maize proteins to be found.

## A glimpse of the maize genome

Several approaches are currently being used to provide further sequence data from the maize genome. These sequences are entered into the Genome Survey Sequence (GSS) division of GenBank because the sequencing is for the most part exploratory, at a low redundancy level. Table 2 summarizes a rough analysis of 11,625 maize GSS entries available as of 1 November 2001. The sequences were obtained by different selection strategies, including genomic sequences flanking Mutator transposon insertions [26], random inserts [27], sequences selected for not being methylated [28], bacterial artificial chromosome (BAC) ends [29], sequences that were genetically mapped [30], and sequences selected for long ORFs using the ORF Rescue vector [31]. Table 2 gives the result of a BLASTP search (option -e 1e-5) of all ORFs of at least 120 codons derived from the GSSs, compared to the non-redundant maize protein set. It can be seen that the random sequencing approaches (random inserts and BAC ends) produce a large fraction of sequences matching

**Table 2**

**Analysis of maize genome survey sequences: a comparison with maize proteins and ESTs**

| Approach | Number of entries | Unique sequences | wORF | Comparison with maize proteins | | | | NS %EST | Reference |
|---|---|---|---|---|---|---|---|---|---|
| | | | | %NS | %TE | %HP | %KP | | |
| Mutator insertions | 4412 | 970 | 375 | 93 | 3 | 2 | 2 | 26 | [26] |
| Random inserts | 3480 | 2529 | 1015 | 61 | 38 | 1 | 1 | 44 | [27] |
| Methylation filter | 1692 | 1083 | 258 | 84 | 10 | 2 | 3 | 37 | [28] |
| BAC ends | 945 | 881 | 454 | 48 | 51 | 0 | 0 | 28 | [29] |
| MPP | 669 | 338 | 150 | 80 | 1 | 7 | 11 | 47 | [30] |
| ORFs | 399 | 86 | 79 | 76 | 0 | 14 | 10 | 22 | [31] |
| Other | 28 | 11 | 3 | 33 | 67 | 0 | 0 | 0 | |

All sequences were retrieved from the GenBank GSS database (with the number of database entries given in the second column). Sequences shorter than 360 bp and redundant sequences were removed with the vmatch program [21] (V.B. and S.K., unpublished), resulting in the reduced sequence set sizes given in the column 'Unique sequences'. Of these, only sequences with non-redundant open reading frames of at least 120 codons (with the number of qualifying entries given in the wORF column) were compared to a maize protein set using BLASTP [23]. Entries were classified on the basis of BLASTP results and GenBank keywords as novel (NS), transposable element (TE), hypothetical protein (HP), or known protein (KP). The corresponding columns give the fraction of sequences in each class (percent). The column 'NS %EST' gives the percentage of sequences with novel ORFs matching maize ESTs. MPP, Missouri Mapping Project.

**Table 3**

**Analysis of maize genome survey sequences: a comparison with *Arabidopsis* proteins**

| Approach | %NS | %TE | %HP | %KP |
|---|---|---|---|---|
| Mutator insertions | 69 | 3 | 24 | 4 |
| Random inserts | 63 | 33 | 3 | 0 |
| Methylation filter | 81 | 7 | 11 | 0 |
| BAC ends | 47 | 48 | 5 | 0 |
| MPP | 59 | 3 | 34 | 4 |
| ORFs | 46 | 1 | 49 | 4 |
| Other | 33 | 67 | 0 | 0 |

GSS sequences with unique ORFs were compared with *Arabidopsis* proteins using BLASTP. Data sets and columns are as in Table 2.

transposable elements, whereas the Mutator transposon insertion, methylation filter, and 'ORF rescue' approaches clearly bias against the recovery of such sequences. More than 80% of the GSS entries with ORFs derived from the former two approaches do not show significant similarities to known maize proteins, and, surprisingly, more than 70% do not match any *Arabidopsis* proteins (Table 3). An intriguing explanation would be that these ORFs correspond to novel or highly diverged maize proteins. It is also possible that some of the ORFs do not correspond to native translation products.

To assess these possibilities, we compared the sequences of novel ORFs with the maize EST set (application of GeneSeqer [22]). The result, that 26-44% of the four large GSS collections match (a still limited collection of) maize ESTs (see Table 2), suggests that many of the ORFs do indeed correspond to expressed genes. The remaining fraction may include less abundantly expressed genes. We can estimate the gene fraction accessible by EST sequencing from the EST coverage of GSS-derived ORFs: if the roughly 10,000 novel ORFs in the maize EST set constitute only 40% of the genes, we can anticipate some 25,000 novel maize proteins that are not found in *Arabidopsis*. It is likely that many of these proteins are derived from gene duplications. The lack of sequence conservation across the monocot-dicot divide suggests that there has been extensive functional divergence after duplication.

## The need for a maize genome sequencing project

On the basis of available data, we think that the resource provided by the *Arabidopsis* genome cannot adequately substitute for more extensive maize genome sequencing. Genome organization is very different between the two plants, and the proteomes may also have significant differences, particularly with respect to agronomically important maize genes involved in plant-pathogen interactions, reproduction, and the development and function of specific tissues. The many exceptions to micro-colinearity even among the grasses suggest that the completion of the rice genome [32] will still not answer many of the questions particular to maize genomics. Beyond questions concerning agronomically important traits, plant biologists also look to maize as a model for the evolution of plant genomes that are not as small and streamlined as those of *Arabidopsis* and rice [33]. Correspondingly, a maize genome sequencing project will focus on sequencing gene-rich genome fractions first [34], and other crop genome projects are likely to follow. Plant biologists should look forward to very exciting times when whole-genome comparisons become possible, leading to a clearer understanding of the development of plants from their genetic blueprints.

## Acknowledgements

V.B. and V.W. were supported in part by NSF Plant Genome Research Program grant DBI-9872657. S.K. was partially supported by grant KU 1257/1 from the Deutsche Forschungsgemeinschaft. The authors are grateful to Phil Becraft, Alan Myers, Tom Peterson, Pat Schnable, and Robert Thornburg for critical comments on the manuscript.

## References

1. Wang R-L, Stec A, Hey J, Lukens L, Doebley J: **The limits of selection during maize domestication.** *Nature* 1999, **398:**236-239.
2. Kellogg EA: **Evolutionary history of the grasses.** *Plant Physiol* 2001, **125:**1198-1205.
3. **Freeling M:** Grasses as a single genetic system. Reassessment 2001. *Plant Physiol* 2001, **125:**1191-1197.
4. Gaut BS, Doebley JF: **DNA sequence evidence for the segmental allotetraploid origin of maize.** *Proc Natl Acad Sci USA* 1997, **94:**6809-6814.
5. White S, Doebley J: **Of genes and genomes and the origin of maize.** *Trends Genet* 1998, **14:**327-332.
6. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408:**796-815.
7. Walbot V: **A green chapter in the book of life.** *Nature* 2000, **408:**794-795.
8. O'Neill CM, Bancroft I: **Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*.** *Plant J* 2000, **23:**233-243.
9. Paterson AH, Lan T, Amasino R, Osborn TC, Quiros C: ***Brassica* genomics: a complement to, and early beneficiary of, the *Arabidopsis* sequence.** *Genome Biol* 2001, **2:**reviews1011.1-1011.4.
10. Grant D, Cregan P, Shoemaker RC: **Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*.** *Proc Natl Acad Sci USA* 2000, **97:**4168-4173.
11. Ku HM, Liu JP, Doganlar S, Tanksley SD: **Exploitation of *Arabidopsis*-tomato synteny to construct a high-resolution map of the ovate-containing region in tomato chromosome 2.** *Genome* 2001, **44:**470-475.
12. Mysore KS, Tuori, RP, Martin GB: ***Arabidopsis* genome sequence as a tool for functional genomics in tomato.** *Genome Biol* 2001, **2:**reviews1003.1-1003.4.
13. Civardi L, Xia YJ, Edwards K, Schnable PS, Nikolau BJ: **The relationship between the genetic and physical distances of the cloned *a1-sh2* interval of the *Zea mays* L. genome.** *Proc Natl Acad Sci USA* 1994, **91:**8268-8272.
14. Chen M, SanMiguel P, De Oliveira AC, Woo S-S, Zhang H, Wing RA, Bennetzen JL: **Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes.** *Proc Natl Acad Sci USA* 1997, **94:**3431-3435.

15. Bennetzen JL, SanMiguel P, Chen MS, Tikhonov A, Francki M, Avramova Z: **Grass genomes.** *Proc Natl Acad Sci USA* 1998, **95:**1975-1978.
16. Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z: **Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum.** *Proc Natl Acad Sci USA* 1999, **96:**7409-7414.
17. Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A: **The complete sequence of 340 kb of DNA around the rice *Adh1-Adh2* region reveals interrupted colinearity with maize chromosome 4.** *Plant Cell* 2000, **12:**381-391.
18. Fu HH, Park WK, Yan XH, Zheng ZW, Shen BZ, Dooner HK: **The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome.** *Proc Natl Acad Sci USA* 2001, **98:**8903-8908.
19. Yao H, Zhou Q, Li J, Smith H, Yandeau M, Nikolau B, Schnable PS: **Meiotic recombination across the 140 kb multigenic maize *a1-sh2* interval.** *Proc Natl Acad Sci USA*, *in press*.
20. Bennetzen JL: **Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions.** *Plant Cell* 2000, **12:**1021-1029.
21. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R: **REPuter: the manifold applications of repeat analysis on a genomic scale.** *Nucleic Acids Res* 2001, **29:**4633-4642.
22. **TIGR FTP site: *Arabidopsis* putative proteins** [ftp://ftp.tigr.org/pub/data/a_thaliana/ath1/SEQUENCES/ATH1.pep]
23, Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.
24. Usuka J, Brendel V: **Gene structure prediction by spliced alignment of genomic DNA with protein sequences: increased accuracy by differential splice site scoring.** *J Mol Biol* 2000, **297:**1075-1085.
25. **dbEST** [http://www.ncbi.nlm.nih.gov/dbEST]
26. **ZmDB: Maize genome database** [http://www.zmdb.iastate.edu/]
27. Meyers BC, Tingey SV, Morgante M: **Abundance, distribution and transcriptional activity of repetitive elements in the maize genome.** *Genome Res* 2001, **11:**1660-1676.
28. Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, Stein L, McCombie WR, Martienssen RA: **Differential methylation of genes and retrotransposons allows shotgun sequencing of the maize genome.** *Nat Genet* 1999, **23:**305-308.
29. **Maize ZMMBBb STC project** [http://www.genome.clemson.edu/projects/stc/maize/ZMMBBb/index.html]
30. **Maize mapping project** [http://www.cafnr.missouri.edu/mmp/]
31. **ISU maize genome project** [http://maize.math.iastate.edu/isumaize/homepage.html]
32. Yu J, Hu S, Wang J, Li S, Wong KG, Liu B, Deng Y, Dal L, Zhou Y, Zhang X, *et al.*: **A draft sequence of the rice (*Oryza sativa* ssp. Indica) genome.** *Chinese Sci Bull* 2001, **46:**1937-1942.
33. Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC: **Maize as a model for the evolution of plant nuclear genomes.** *Proc Natl Acad Sci USA* 2000, **97:**7008-7015.
34. Bennetzen JL, Chandler VL, Schnable P: **National Science Foundation-sponsored workshop report. Maize genome sequencing project.** *Plant Phys* 2001, **127:**1572-1578.